Exploring linkages between weather factors and the risk of cerebral infarction through the application of Bayesian networks

Hiroshi Morimoto

Abstract— **A number of studies, mainly applying regression models, have investigated the correlation between meteorological factors and the incidence of cerebral infarction. One assumed subjectively the roles of the variables of the onset of cerebral infarction and the variables of the change in weather, and tried to find correlation. However, correlation can not induce causality. An exploratory method such as the Bayesian network approach, which is based on exhaustive searches, is thus required to determine causality. This approach considers all the possible causal associations and determines the most probable causality by estimating network scores. This study was based on the daily data on the number of patients transported by ambulance in the city of Nagoya in Japan. The patients were first transported to hospitals by ambulance prior to being diagnosed with cerebral infarction. The use of heuristic search techniques enabled us to determine a causal relation between the weather and the incidence of cerebral infarction, including the influence of weather states such as high pressure or low pressure types considered as one of the nodes of the network. At the same time, we observed the influence of delayed effects of cold exposure on the onset of cerebral infarction. As an application of these findings, we showed the existence of the threshold for the mean temperature that divided risky temperature and safe temperature. These results provided clear evidence of the links between human physiological responses and changes in weather, and suggest the possibility of predicting cerebral infarctions.**

Index Terms—**Bayesian network, cerebral infarction, stroke, weather.**

I. Introduction

While a number of studies, using standard statistical techniques such as regression analysis (e.g., [1–2]) have investigated the correlation between meteorological factors and the incidence of cerebral infarction, their links remain unclear. In this paper, we present a new approach for addressing this issue using Bayesian networks. A Bayesian network is a graphical model for representing conditional independencies within a set of random variables. A Bayesian network inference algorithms have shown particular promise, because unlike standard regression models, they can capture more realistic relationships between variables, because of their probabilistic nature [3]. Here, we applied Bayesian network framework and aimed at evaluating and understanding the links between environmental conditions and diseases.

In this paper, the simplest type of a Bayesian network entailing three nodes was considered. The first node represented the risk of cerebral infarction. The second node denoted meteorological factors such as the mean temperature. The third node expressed the background state of the weather. The importance of the third node has already been demonstrated by [4], with respect to hidden Markov model (HMM).

When applying a Bayesian network, the basic question is whether an arrow can be drawn from one node to another node. A heuristic algorithm, based on a Bayesian score, was used to determine this. For example, if an arrow could be drawn from the second node of the meteorological element to the first node of the risk of cerebral infarction, this indicated a causal effect extending from the second node to the first node.

The general approach used was quite simple. We applied the heuristic method of the Bayesian network to assess the links between the effects of weather and the incidence of cerebral infarction. Once we had established the Bayesian network, we were able to apply probabilistic inference techniques. For example, we posit a temperature threshold separating risky from safe temperatures, based on a probabilistic perspective. Here, the threshold indicated that a drop in temperature below a certain level (i.e., a threshold) in winter was directly associated with an increased incidence of cerebral infarction. In summer, the direction of these effects was reversed.

II. Methods

This study was based on the data of daily number of patients in 2003, obtained from the city of Nagoya. Nagoya is located in central Japan, facing the Pacific Ocean. It enjoys a typical mild Japanese climate that is clearly associated with the four characteristic annual seasons. The data included the number of patients of all ages, who were first transported by ambulance to a hospital and then diagnosed, at the hospital, with cerebral infarction. The meteorological data comprised a selection of daily data that was obtained from the Japan Meteorological Agency. These data included temperature (mean, maximum and minimum temperatures) and the hours of sunshine and so on. We also used previously classified daily data on weather states [4]. The classification of these states was carried out using self-organized mappings.

The following six groups of weather states were identified: "a warm type of high air pressure air", "a cold type of high air pressure air", "a warm type of low air pressure air", "a cold type of low air pressure air", " a humid type of rain" and " a type of cold wind". These categories were effectively expressed the background states for the incidence of cerebral infarction [4].

Hiroshi Morimoto, Graduate School of Environmental Study, Nagoya University, Nagoya, Japan, 81-52-789-4221

Bayesian networks have emerged as powerful computational tools for identifying putative causal interactions among variables from observational data. Bayesian network inference algorithms hold particular promise, as they can capture linear, non-linear, combinatorial, stochastic and other types of relationships across multiple levels of variables [5].

Bayesian network analysis is based on a learning approach for addressing the problem of optimization. It starts by defining a statistically motivated "score" that describes the fit of each possible structure to the observed data. During each step of the learning process, a score is computed for each network *G*. A scoring metric enables us to evaluate how probable it is that the network can explain relationships among observed data *D*. We conducted a heuristic search to identify the network with the highest Bayesian score [6]. The problem of optimization was solved by performing a learning procedure. The procedure is to examine all possible local changes in each step and then to apply the one that leads to the biggest improvement in score. The choice for local changes includes edge addition, edge deletion, and edge reversal.

The probability P(*G*|*D*) denoted the probability of *G* based on the assumption of the provision of data *D*. This probability is known as Bayesian posterior probability and constitutes the basis of Bayesian scores . The classical formulation of Bayes's theorem asserts the following:

$$P(G\,|\,D)=\frac{P(D\,|\,G)P(G)}{\sum_{G_i}P(D\,|\,G)P(G_i)}$$

Usually the denominator of the equation expressing Bayes's Theorem is difficult to calculate. However, what is requested is the one for comparing the two networks G and G'. Therefore the following formula is normally used for obtaining score:

$$\frac{P(D\,|\,G)}{P(D\,|\,G')}$$

Known as the Bayes factor, this formula has practical application. In this paper, *D* denotes the collection of three types of data: meteorological data, risks of cerebral infarction and global weather patterns.

III. Results

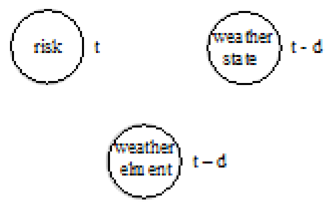
A. Lagged Bayesian network

Here, we outline the framework of our Bayesian networks. We considered three nodes, namely, "risk", "meteorological data" (e.g., mean temperature) and "weather state". The first node "risk" comprised the daily data on the number of patients who were transported by ambulance to a hospital and diagnosed with cerebral infarction. The second node " meteorological data" comprised daily data on mean temperatures in Nagoya. The third node "weather state" comprised daily data describing a type of weather chart or weather class. A previous study [7] introduced and used them as background weather states in order to construct hidden Markov models (HMMs). Self-organized maps were used for the classification. The classified weather patterns comprised the six previously described patterns.

We applied a heuristic learning algorithm of Bayesian network to the three nodes, by calculating network scores and comparing scores, thereby obtaining networks. During each step of the process of developing Bayesian networks, a series of networks was created through addition, deletion or reversal of an arrow. The network scores of all of the proposed networks were calculated and the network with the highest score was chosen for the next step in the search. If none of the proposed networks had a higher network score than the previous network, the search was terminated. Thus we determined the highest network score. For this purpose, we used the "R" software and its "deal" package.

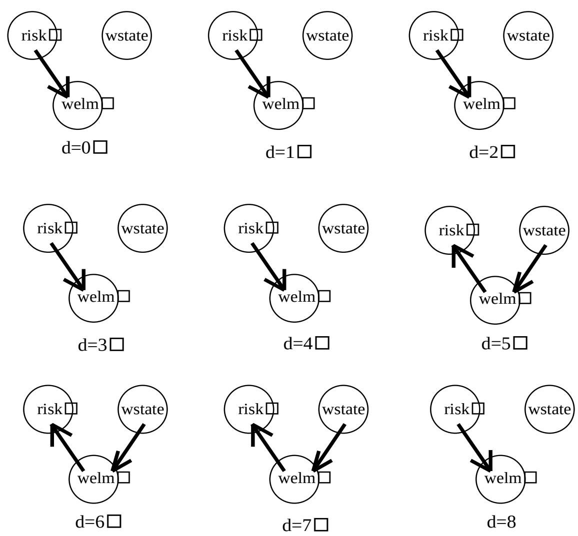
We first presumed that no delay had occurred. In other words, the mean day temperature and the weather pattern affected on the onset of cerebral infarction on the same day. In this case we failed to obtain an effective network. Next, fixing integer *d*, we shifted the time (or, more precisely, "day") , assuming that the mean temperature at time *t-d* and the weather pattern prevailing at *t-d* affected on the onset of cerebral infarction at time *t*. If there was no shift or delay, this implied that *d*=0. Therefore *d* denoted a "delay" or "lag" in this study. See Fig. 1, where two nodes other than "risk node" had lagged time. We termed a network emerging from these nodes, a "Lagged Bayesian network". The network could be thought as one of dynamic Bayesian network.

Figure 1. The three nodes of the lagged Bayesian network with lag = *d* (days).



Note: Here, "weather states" means weather patterns or charts such as "lower air pressure and cold pattern" and so on. "Weather element" denotes meteorological factors such as mean temperature.

Figure 2. Lagged Bayesian Networks with lag from 0 to 8.



Note: Here, "welm" denoted "weather elements" and "wstate" denoted "weather state" like Fig. 1.

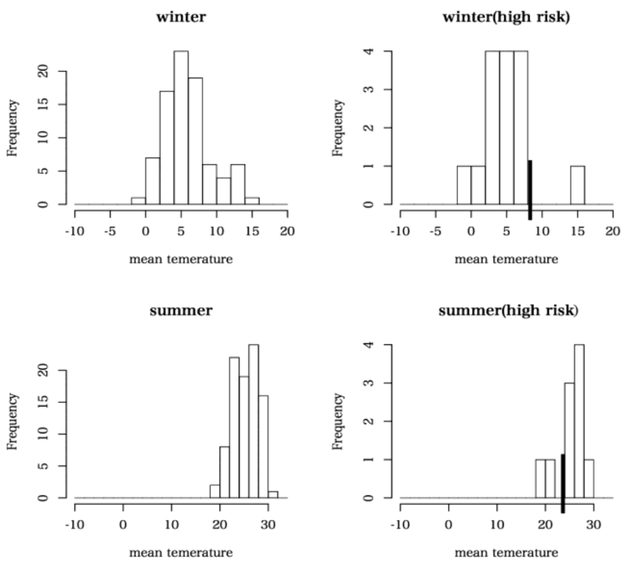
By allowing a variation in values of delay *d*, ranging from 0 to 8, we obtained nine patterns of "lagged Bayesian networks". These networks are depicted in Fig. 2. The network located in the top-left corner of the figure had no lag. The network located in the bottom-right corner had a lag where *d*=8. The most natural networks in Fig. 2 were those with lag entailing a d value of 5,6 and 7. This finding implies that there was a plausible link between weather and disease, and that a delay (*d*=5,6 and 7) existed in relation to the onset of cerebral infarction after the occurrence of a change in the weather.

B. *Threshold for mean temperature*

Having established Bayesian networks with delay values of *d*=5,6 and 7, we fixed a value of *d*=6 to test the application of a Bayesian network. Adopting a probabilistic perspective, we attempted to determine the temperature threshold separating risky from safe temperatures.

The number of patients (per day) who were transported to hospitals in 2003 varied from 0 to 13. We classified these daily data into three categories: "high risk cases", "middle risk cases" and "low risk cases". The classification was performed so that 20 % of the total cases became "high risk", 60% became "middle risk" and 20% became "low risk". We then compared two distributions of the mean temperature. One was the distribution of mean temperature in 2003. This distribution was called the distribution with "no evidence". The other was the distribution of mean temperature of high risk cases. This distribution was called the distribution with "evidence" in the frame work of Bayesian network. The comparison was performed for winter and summer in 2003, respectively. These distributions were shown in Fig. 3. The thick lines in the figure indicate the thresholds for risky mean temperature for both winter and summer in 2003. In winter, the risk was observed to increase if the temperature dropped below 7 or 8 degrees Celsius. In summer, the risk increased if the temperature rose above 24 degrees Celsius.

Figure 3. The distributions of mean temperature.



Note: Here, the thick lines denoted the thresholds separating risky from safe temperature zones.

IV. Discussion

In this study, we constructed a Bayesian network connecting the risks of cerebral infarction, meteorological variables and weather states. The method was based on an exhaustive and explorative search.

We found that the lagged Bayesian network was effective to connect weather and human health.

The shifting method enabled us to evaluate the delay-effect of sudden weather changes on human health. It further revealed links between weather changes and the onset of cerebral infarction.

Out of the available Bayesian network nodes, we selected the node of "weather states". These weather states were derived from the classification of weather charts, and played an important role in Hidden Markov Models (HMMs) in [8]. The classification of weather states was carried out using self-organized maps. These classified weather states were regarded as "hidden" or "background states". The "observed values" in HMM were the risk of cerebral infarction. From a Bayesian network perspective, hidden Markov models represent a special case of dynamic Bayesian network entailing two nodes (i.e., a node of risk of cerebral infarction and a node of weather state). Therefore, our Bayesian network could be regarded as a natural extension of hidden Markov models in [7-8].

An advantage of using a Bayesian network was that it enabled us to locate the thresholds for risky mean temperatures during the winter and summer of 2003 that caused incidents of cerebral infarction as shown in Fig. 3.The study revealed that the risk of cerebral infarction increased in winter when the temperature dropped below 7 or 8 degrees Celsius. During summer, the risk increased when the temperature rose above 24 degrees Celsius. These findings reveal the effectiveness of the probabilistic method which is associated with Bayesian network analysis, and its potential use in forecasting the onset of cerebral infarction.

We observed the occurrence of lags for the effects of weather changes in relation to the onset of cerebral infarction. The existence of lag periods could be indicative of the mechanism whereby weather changes influence the onset of diseases. Therefore, responses to cold exposure become an important consideration.

First, exposure to cold results in increased blood pressure, especially in arterial blood pressure, as an immediate response [9-10]. Second, the increase in blood pressure causes increases in platelet counts and neutrophil counts, and in fibrinogen as a physiological response [10]. During a period of several days following exposure to cold, the increases in platelet counts and fibrinogen deduce platelet viscosity and hemoconcentration, and promote of thrombogenesis [11]. The initiation of a mild inflammatory response has also been reported [11].

At a molecular level, immunohistochemical analysis has revealed that the cold exposure induces vascularization [12]. It has also been reported that exposure to cold results in extensive changes in the expression of genes involved in glycerolipid metabolism and fatty acid elongation [13]. These studies conducted on adipose tissue, revealed the importance

of the mechanism associated with blood vessels for responding to cold exposure.

These findings suggest that onset of cerebral infarction occurs several days after exposure to cold. The mechanism of the lagged effects is highly significant and should be further discussed.

V. CONCLUSION

The application of Bayesian networks was found to be effective in revealing links between weather changes and incidence of diseases. The use of heuristic algorithms associated with this method extends beyond standard regression models, and lead to potential advances. The application of a Bayesian network approach clearly revealed the causal relation between weather changes and cerebral infarction. Furthermore, it confirmed the lagged effect of weather changes and enabled the identification of the threshold that constitutes the boundary between risky and safe temperature.

ACKNOWLEDGMENT

We thank the Nagoya City Fire Department for the offer of the daily data of ambulance transport in the city of Nagoya.

REFERENCES

[1] G,C. Donaldson. and W.R.Keatinge, "Early increases in ischaemic heart disease mortality dissociated from and later changes associated with respiratory mortality after cold weather in south east England" Journal of Epidemiology and Community Health, 1997, 51, 643-648. http://dx.doi.org/10.1136/jech.51.6.643

[2] J. Jimenez-Conde, A. Ois , M.Gomis, A.Rodriguez-Campello, E.Cuadrado-Godia, I.Subirana and J.Roquer , "Weather as a Trigger of Stroke, Daily Meteorological Factors and Incidence of Stroke Subtypes" Cerebrovascular Diseases, 2008, Vol. 26, 348-354. http://dx.doi.org/10.1159/000151637

[3] N.Friedman, I. Nachmanand D. Peér, "Learning bayesian network structure from massive datasets: the «sparse candidate« algorithm" UAI'99 Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence, 1999, Pages 206-215. Morgan Kaufmann Publishers Inc. San Francisco, ISBN:1-55860-614-9

[4] H. Morimoto, "Hidden Markov Models and Self-Organizing Maps Applied to Stroke Incidence" Open Journal of Applied Sciences, 2016, 6,158-168. http://dx.doi.org/http://dx.doi.org/10.4236/ojapps.2016.63017

[5] J. Yu, V.A. Smith, P.P.Wang, A.J. Hartemink and E.D. Jarvis, "Advances to Bayesian network inference for generating causal networks from observational biological data" Bioinformatics , 2004, 20 (18), 3594-3603. http://dx.doi.org/10.1093/bioinformatics/bth448

[6] D.Heckerman, D. Geiger and D.M. Chickering, "Learning Bayesian networks: The combination of knowledge and statistical data" Machine Learning, 1995, 20:197-243.http://dx.doi.org/10.1007/BF00994016

[7] H. Morimoto, "Study on Links between Cerebral Infarction and Climate Change Based on Hidden Markov Models" International Journal of Social Science Studies, 2015, 3, 180-186. http://dx.doi.org/10.11114/ijsss.v3i5.1045.

[8] H. Morimoto, "Hidden Markov models to estimate the lagged effects of weather on stroke and ischemic heart disease" Applied Mathematics, 2016 vol.7, 1415-1425. http://dx.doi.org/10.4236/am.2016.71312

[9] F. Ballester, D, Corella, S. Pperez-Hoyos, M. Saez and A. Hervas, "Mortality as a Function of Temperature. A Study in Valencia, Spain, 1991–1993" International Journal of Epidemiology, 1997, Vol. 26, No. 3, 551-561. http://dx.doi.org/10.1093/ije/26.3.551

[10] P.C Eldwood, A. Beswick, J. R.O'Brien, S.Renaud, R .Fifield, E. S.Limb, D.Bainton, "Temperature and risk factors for ischaemic heart disease in the Caerphilly prospective study" Br Heart J , 1993,1993;70:520-523. http://dx.doi.org/10.1136/hrt.70.6.520

[11] J.B.Mercer, "Cold—an underrated risk factor for health" Environmental Research, 2003, 92, 8–13. doi:http://dx.doi.org/10.1016/S0013-9351(02)00009-9

[12] Y.Xue, N.Petrovic, R.Cao, O.Larsson, S.,Lim, S.Chen, H.M.,Feldmann, Z.Liang, Z.Zhu, J.Nedergaard, B.Cannon and Y. Cao, "Hypoxia-Independent Angiogenesis in Adipose Tissues during Cold Acclimation" Cell Metabolism, 2009, 9, 99–109. http://dx.doi.org/10.1016/j.cmet.2008.11.009

[13] A.B.Marcher, A.Loft, R.Nielsen, M.Sysi-Aho, K.Ekroos, S.Mandrup, "RNA-Seq and Mass-Spectrometry-Based Lipidomics Reveal Extensive Changes of Glycerolipid Pathways in Brown Adipose Tissue in Response to Cold" Cell Reports, 2015, 13, 2000–2013. http://dx.doi.org/10.1016/j.celrep.2015.10.069

Hiroshi Morimoto Dr.Hiroshi Morimoto was graduated from Mathematical Institute of Nagoya University. He was first employed by Nagoya University as a pure Mathematician. Then he explored many fields including the field of global climate change and the analysis of big data. Now he is majoring the field of bioweather, working as a professor of the same university.